

PJM Potential Load Forecast Changes

Questions, Concerns and Comments

Stephen Wreschnig

Manager, Electric and Gas Sales and Revenue Forecasting

Public Service Electric & Gas

August 15, 2019



Introduction

PJM has requested that the members of the Load Analysis Subcommittee (LAS) and any stakeholders examine the documents provided for the August 6, 2019 LAS meeting describing proposed load forecast model changes.

After reviewing the materials, I feel that the proposed model by PJM is not ready to be put into production. The questions and issues that the methodology and data utilization raise are serious. These questions are compounded by the fact that the accuracy of the estimation of a key component, non-weather sensitive load, cannot be determined. As a result, I feel that these questions and issues need to be resolved before any further steps are taken.

My questions, comments, and concerns follow. On the next slide, I summarize my thoughts on the various changes that have been suggested and note the slides that contain more detail.

I'm very interested in hearing the opinions of other LAS members and stakeholders in addition to the PJM staff so that a comprehensive discussion of these model changes will follow. This type of discussion is very difficult in the LAS meeting/conference call environment when a great deal of material is presented and the issues are complex.

Summary

Change # 1 – End-Use Characteristics

Slides 3-5

- Weighting of the Equipment Indices based on annual sales by customer sector would impact the validity of the peak load estimations given varying sector response to weather, and need clarity on some of the non-residential energy drivers.

Change # 2 – Modeling Non-weather Sensitive Load

Slides 6-9

- The two-stage least squares is used inappropriately to estimate base use utilizing equations that, as a result, suffer from omitted variable bias. Also, the limited selection of months and use of yearly binaries is questionable. Both will have a major impact on the validity of the peak load estimations. Additionally, the use of the Other Equipment Index is not clear.

Change # 3 – Modeling Non-weather Sensitive Load

Slides 10-12

- Relying on CDD fails to take humidity into account, and the structure of some of the lag values is questionable. Also, the weighting process is flawed from an omitted variable bias perspective and from a structural form that's inconsistent from the assumed relationship of weather and peak in the peak model.

Change # 4 – Distributed Solar Treatment

Slide 13

- Estimating the relationship between load and solar output should be placed on the list of future model enhancements to enable a better matching of solar capacity to hourly loads.

Change # 5 – Explicit treatment of Plug-in Electric Vehicles

Slide 14

- I know of two sources of electric vehicle data that may be useful.

PJM RTO Energy Forecast

Slides 15-16

- The existing model has errors that are constantly increasing while the proposed model has a errors that are constantly decreasing and, I would expect, become increasingly negative over time. The change in the RTO proposed forecast is consistent with a model whose errors are becoming increasingly negative over time.

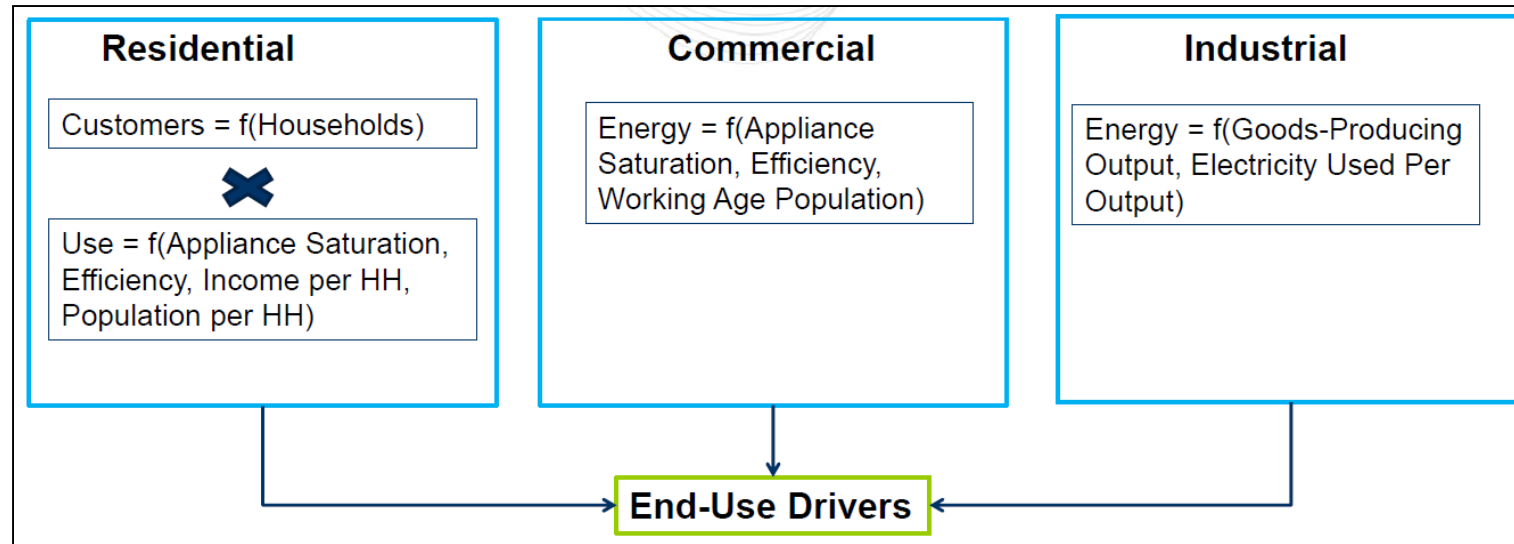
Change # 1 – End-Use Characteristics

- Current
 - Equipment indexes are computed for Residential and Commercial sectors based on information from the EIA/Itron. These are then weighted according to Residential and Commercial energy sales from EIA 861 for the last 5 years.
- Proposed
 - Further leverage EIA 861 data. Use to calibrate Equipment Indexes. Historic and forecast weights of different sectors vary over time.

PJM, "Potential Load Forecast Changes", P. 12

- My understanding is that the Equipment Indices should be efficiency-weighted measures of the appliance stock.
- Weighting the indices annually based on energy sales on an annual basis will cause the aggregate index to fluctuate based on the weather since different sectors have different ranges of weather response.
- This would result in biased estimates of the estimated coefficients in the peak load model as the weather impact and equipment indices are misstated.

Change # 1 – Using EIA 861 Data

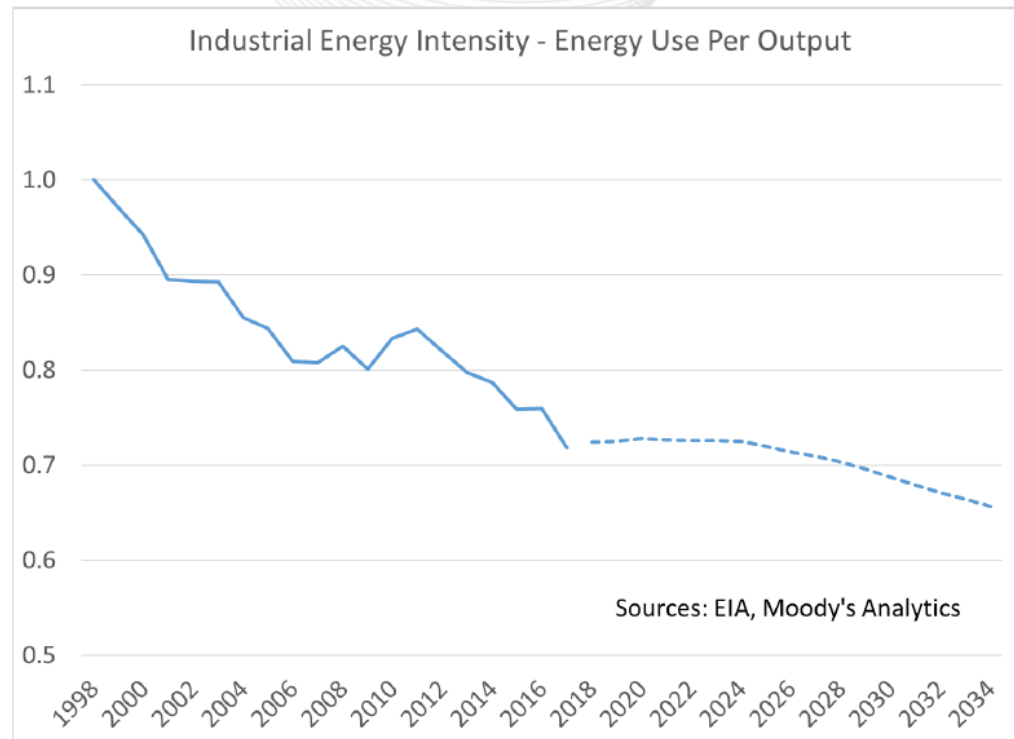


PJM, "Potential Load Forecast Changes", P. 13

- What is the "Working Age Population" in the Commercial drivers supposed to represent? It doesn't seem to be a good measure of commercial employment/activity or of the demand for commercial services. Income-based and population-based measures seem to be better drivers of the demand for commercial services.
- I'm a little confused by the Industrial specification. It appears to be a tautology where:
$$\text{Energy} = f(\text{Goods-Producing Output, electricity Used Per Output}) = \text{Energy}$$

I think that I need some clarification of what you're trying to do here.

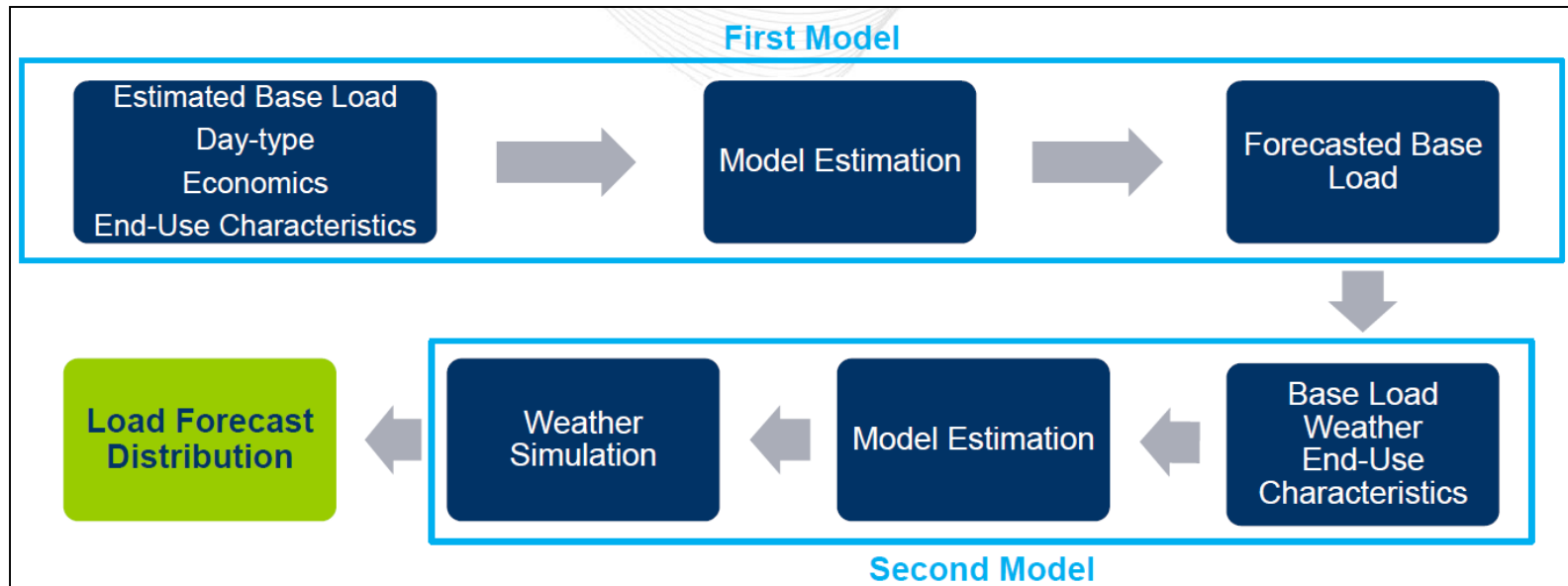
Change # 1 – Considering Industrial Intensity



PJM, "Potential Load Forecast Changes", P. 14

- What is Moody's rationale for the abrupt change in the trend that has been seen in the Energy Use Per Output that occurs during the forecast period?

Change # 2 – Modeling Non-weather Sensitive Load



PJM, "Potential Load Forecast Changes", P. 11

- I have concerns about the two-stage least squares (2SLS) type estimation process that you're employing.
- It doesn't appear that you're employing 2SLS to create an instrumental variable or to deal with an over identified structural equation that's part of a system of equations – two of the major reasons for using 2SLS.

Change # 2 – Modeling Non-weather Sensitive Load (continued)

- It appears that you're trying to use your two regressions to mimic a technique widely used in natural gas consumption to estimate base use for many years.
 - With gas, the consumption in the summer months is used as an estimate of base load.
 - This works well with gas because the major use for gas is space heating that doesn't occur in the summer.
 - The gas consumption of secondary uses, water heating and cooking, have much lower correlation with weather.
- This technique is, however, much more problematic when dealing with electricity.
 - Electricity, unlike gas, is sensitive to both winter and summer weather.
 - There are no months in the year that have no weather component of electricity consumption
 - There are numerous end-uses for electricity – each of which can have their own unique seasonal patterns that may or may not be weather related.

Change # 2 – Modeling Non-weather Sensitive Load (continued)

- Calculated by modeling shoulder month load (Mar-May; Sep-Nov) and accounting for heating and cooling, growth, and weekends

Load = f(Heating, Cooling, Yearly Binaries, Weekend Binaries)

Non-weather sensitive load = Yearly Binaries

PJM, "Potential Load Forecast Changes", P. 4

- Limiting your sample to the shoulder months is assuming that the non-weather sensitive load in the other months is a constant multiple of the shoulder month baseload across time (since it is not incorporated into the peak equation with a time varying parameter). This does not seem to be a valid assumption to me and a source of bias in your peak model parameter estimates.
- It appears that the only variable capturing the trend in the non-weather sensitive load over time is the Yearly Binaries. Using binary variables as the only variables to capture annual changes results in the annual binaries capturing all of the omitted variables, such as changing customer class distribution, any imprecision of the weather variables, or any other phenomena. Modeling non-weather sensitive load over time with no explanatory variables appears to be a major shortcoming particularly in a forecast model where changes over time are key.

Change # 2 – Modeling Non-weather Sensitive Load (continued)

- Proposed
 - Estimate non-weather sensitive load. Forecast using Other Equipment Index and use this as an input to the model.

PJM, “Potential Load Forecast Changes”, P. 16

- If the Other Equipment Index isn't in the non-weather sensitive model specification (per PJM, “Potential Load Forecast Changes”, P. 4), how are you forecasting it with the Other Equipment Index?
- If the Other Equipment Index is a rational variable to forecast the non-weather sensitive load, it needs to be part of the non-weather sensitive load estimation equation. That is the manner in which its contribution to the growth of non-weather sensitive load can be quantified and utilized in the forecast.

Change # 3 – Building Weather Variables

- Weather variables used:
 - Summer – CDD, Lag CDD, max daily THI
 - Winter - 3 Hr MA WWP at time of peak, Average Daily WWP, Lag Average Daily WWP, and Minimum Daily WWP

PJM, "Potential Load Forecast Changes", P. 20

- Why are you using both CDD and THI? The general consensus is that cooling load is a function of temperature and humidity so what's the justification of using CDD, a temperature-based measure?
- Has it been verified that the max daily THI never occurs AFTER the peak?
- What is the basis for the variables that are used in the winter weather variable calculation. Has it been verified that the Minimum Daily WWP never occurs AFTER the peak?
- Why is the average daily WWP variable in the specification when its value may have been significantly influenced by weather that occurred AFTER the peak?

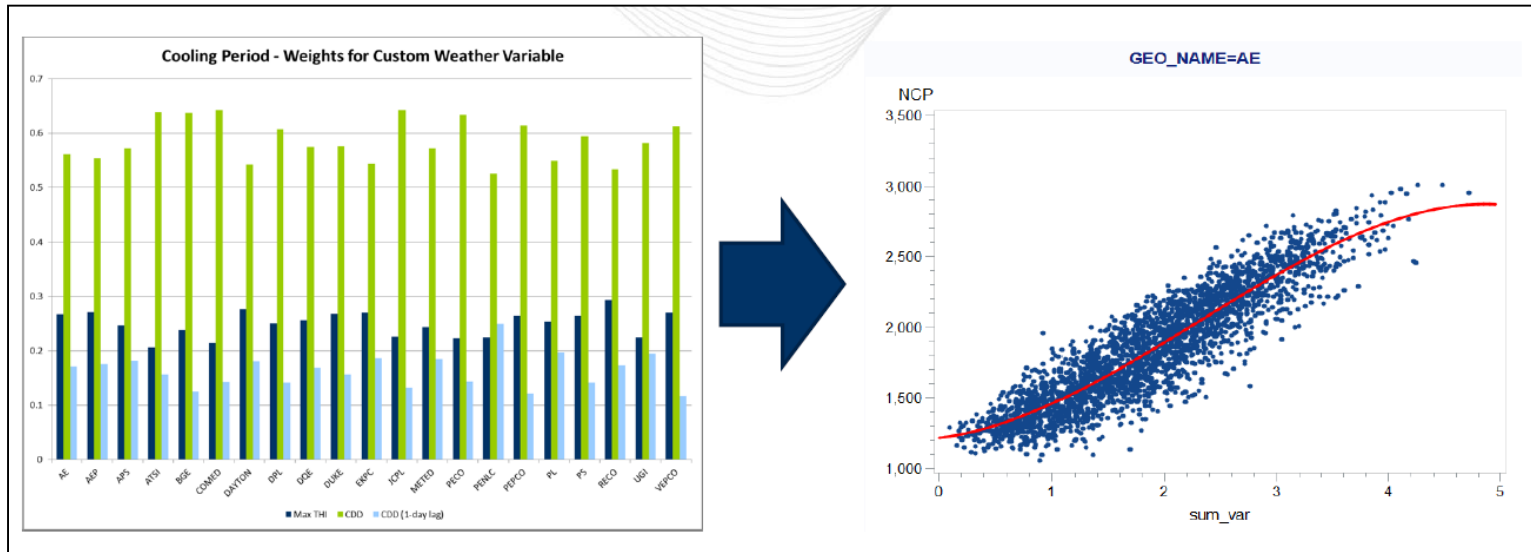
Change # 3 – Building Weather Variables(continued)

- Seasonal regression models are run on load for each weather variable. Sum of Squared Errors (SSE) is saved and then the reciprocals are used as the weights (i.e. the less the error, the higher the weight). The end result is a single weather variable for each season that is a weighted composite.

PJM, “Potential Load Forecast Changes”, P. 21

- Why use separate regressions when your results will all be affected by omitted variable bias. Multiple regression is designed to do what you are doing without this issue. In the presence of multicollinearity multiple regression results may be imprecise but not biased. Your regressions produce biased results.
- A further source of bias resides in the equation specifications. If the weather impacts are non-linear, as stated on pages 23 and 25, what is the justification for using a linear function for the weights?
- The weather variables may be considered the most crucial variables in the peak model yet they are relegated to being summarized in a questionable manner. I don't see a justification for this. While saving degrees of freedom in the estimation process makes sense when there are fifty independent variables in the equation, it might be better to combine some of the numerous binary/dummy variables that do not differ significantly. “Wednesday_Base” and “Thursday_Base” seem to be two likely candidates.

Change # 3 – Building Weather Variables(continued)



PJM, "Potential Load Forecast Changes", P. 22

- From the bar chart it appears that the variable with the greatest weight is CDD. It makes no sense, from a modeling perspective, to have the major summer variable be a temperature-only based variable when humidity plays a key role in air conditioning load.
- What is the evidence supporting using a one day lag for CDD? My experience is that the previous overnight weather is the main lagged driver since it captures potential thermal buildup.
- I don't see two inflection points in the graph. I see a red line that understates NCP in the 4-5 range and overstates the NCP in the 0-1 range. What is the evidence supporting the use of a cubic polynomial?

Change # 4 – Distributed Solar Treatment

- Current
 - Calculate an average capacity factor at time of peak by season (e.g. HE17 in the Summer) . Multiply this by installed capacity to get daily peak reduction attributed to solar. Create gross load peak distribution and then reduce by solar.
- Proposed
 - Calculate historical daily capacity factors. Multiply this by installed capacity and reduce gross load in the forecast simulation.

PJM, "Potential Load Forecast Changes", P. 27

- Using average daily NEM capacity factors will, most likely, still understate the impact of solar since the peak hour will be a time of higher than average load and higher than average solar generation.
- Perhaps trying to generate the relationship between load and solar output could be placed on the list of future model enhancements to enable a better matching of solar capacity to hourly loads.

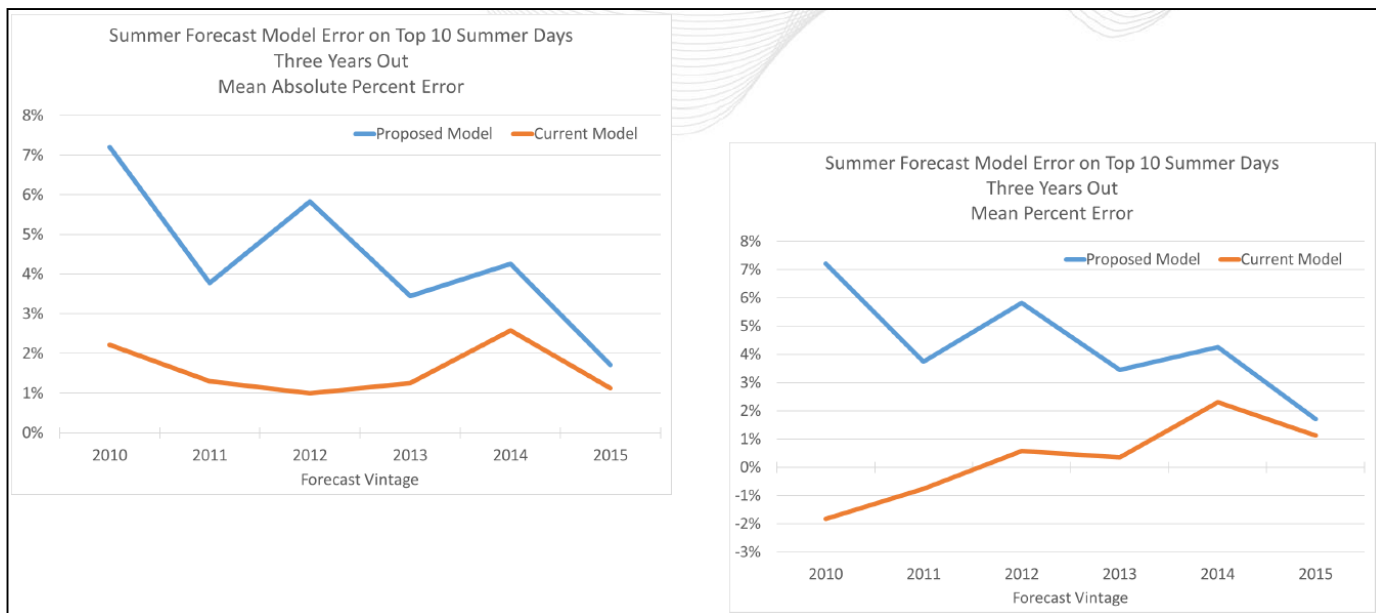
Change # 5 – Explicit treatment of Plug-in Electric Vehicles

- Proposed
 - Explicit treatment of plug-in electric vehicles. Would make assumptions about type and time of charging and add to forecasted load. Preliminary adjustment included on later slides. Still a work in progress, additional presentation on the topic.

PJM, "Potential Load Forecast Changes", P. 29

- A good source of national data is Motorsport Network's web site InsideEVs.com. <https://insideevs.com/news/362819/ev-sales-scorecard-july-2019/> They publish monthly sales totals by make and model and have a listing of vehicle characteristics. It's not downloadable but, I have been keeping track and have a SAS dataset that I can provide if you wish.
- A source of data by state is from the Alliance of Automobile Manufacturers. <https://autoalliance.org/energy-environment/advanced-technology-vehicle-sales-dashboard/> This data is only disaggregated by ATV category, Fuel Cell Vehicles, Battery Electric Vehicles, Plug-in Hybrid Vehicles, and Hybrid Vehicles. It, again is not downloadable and very awkward to obtain. I've just kept track of New Jersey which I can provide if you wish.

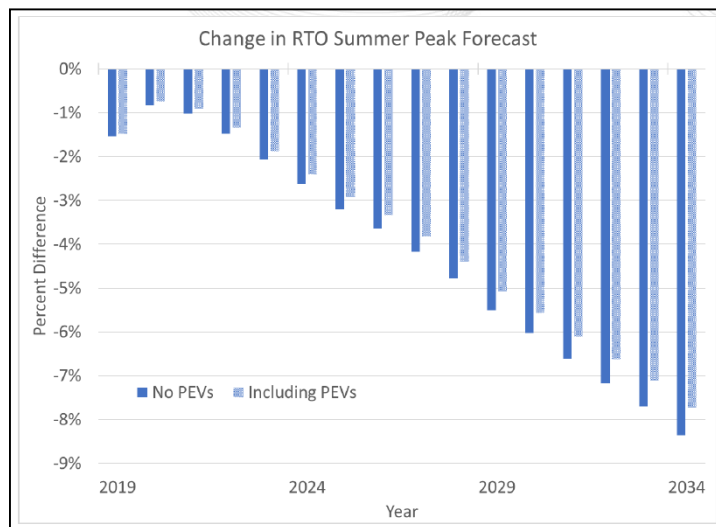
PJM RTO Energy Forecast



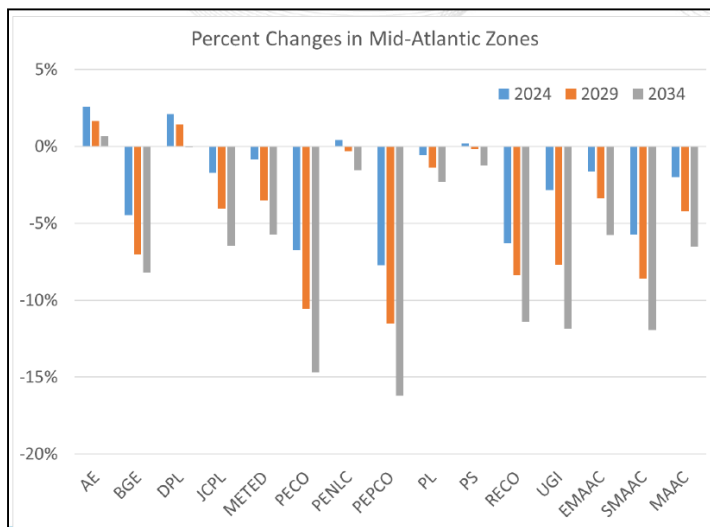
PJM, "Potential Load Forecast Changes", P. 48

- I glean from these graphs that the existing model has errors that are constantly increasing while the proposed model has a errors that are constantly decreasing and, I would expect, become increasingly negative over time.
- I don't see an improvement in the forecasting capabilities, just a change in the direction of the bias of the models.

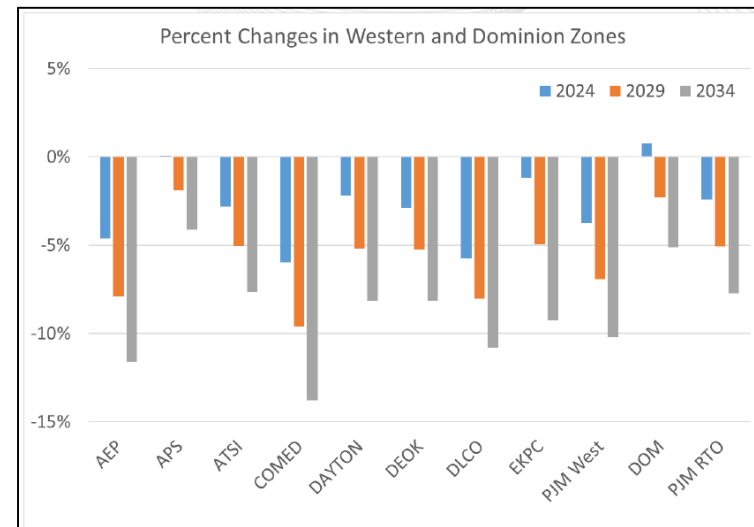
PJM RTO Energy Forecast (continued)



PJM, "Potential Load Forecast Changes", P. 35



PJM, "Potential Load Forecast Changes", P. 37



PJM, "Potential Load Forecast Changes", P. 38

- The change in the RTO forecast is consistent with a model whose errors are becoming increasingly negative over time.
- If it is assumed that it is better to overestimate capacity requirements than underestimate capacity requirements, given the lead times for construction, adopting the proposed model seems very risky.